113-V1 18191

Introduction to Statistics
Dr Matthew Sunderland

[handwritten: matthew.sunderland @ csi.cuny.edu]
[handwritten: Matt]

1. Synchronous lecture Monday Wednesday 12:20–2:15
   https://zoom.us/meeting/register/tJErfuyvrjovHdI9iJLZOZDyB7r6UFMbfomA

2. Online problem sets due Sundays *[handwritten: webwork]*
   https://www.math.csi.cuny.edu/webwork2/Math113_18191_Sunderland_F20/
   Username and pw are both set to your **CUNY username**, eg, first.last00 (*not* jsmith5678)

3. Written assignments due some Sundays on
   https://www.gradescope.com course code M4BW6G

4. Reading assignments due each night before lecture
   https://www.perusall.com course code SUNDERLAND-TSH2K

5. Office hours TBA *[handwritten: we'll use The Openstax Stats text]*
   https://zoom.us/my/mattsunderland

6. Announcements, Lecture Recordings, and Grades posted on
   https://bbhosted.cuny.edu

7. Platform for administering exams TBA,
   possibly Blackboard, Gradescope, WeBWorK, Respondus, or Proctortrack

Day 1 Homework

1. Download Zoom and create free account ✔
2. Do Online Problem Set 1 by Sunday 8/30
3. Submit Written Assignment 1 by Sunday 8/30—see last two pages of syllabus
4. Do first reading assignment (1.1–1.2) and make 2 comments by Sunday 8/30
5. Do office hour survey https://forms.gle/RRf74atLQkR3kg5DA

Course Grade = Average of
- Coursework
  1. Lecture participation
  2. Online problem sets
  3. Written assignments
  4. Reading assignments
- Exam 1
- Exam 2
- Final

[handwritten: zoom, webwork, gradescope, perusall]

Lecture Recording Statement *Students who participate in this class with their camera on or use a profile image are agreeing to have their video or image recorded solely for the purpose of creating a record for students enrolled in the class to refer to, including those enrolled students who are unable to attend live. If you are unwilling to consent to have your profile or video image recorded, be sure to keep your camera off and do not use a profile image. Likewise, students who un-mute during class and participate orally are agreeing to have their voices recorded. If you are not willing to consent to have your voice recorded during class, you will need to keep your mute button activated and communicate exclusively using the "chat" feature, which allows students to type questions and comments live.*

Deadlines   Add 9/1   Drop 9/15   Withdraw 11/6

Written Assignment 1

Name _____   EMPLID _____

Course _____   Date _____

Directions: Upload your completed assignment to Gradescope as a PDF. For full credit, each page of your submission must be right side up and the pages must be in the correct order. If Gradescope asks you to match questions to pages, do so. Many students find it easier to type/annotate directly onto the PDF on the computer; other students prefer to print out the assignment, handwrite their answers, and then use a scanning app to get the completed assignment back onto the computer. If you scan, make sure you scan as a single PDF (with two pages) and make sure you scan as a document, not a picture (completely white background between text).

e.g. dropbox app
Scannable
Scanner app

Question 1.

Is this your first math course at CSI? If not, what math course did you take before this?

Question 2.

What is your preferred email address?

Question 3.

Are you taking this course to satisfy a requirement? Some other reason?

Question 4.

Is there some grade in this course that you will strive to make?

stockbroker uses probability to determine the rate of return on a client's investments. You might use probability to decide to buy a lottery ticket or not. In your study of statistics, you will use the power of mathematics through probability calculations to analyze and interpret your data.

Key Terms

In statistics, we generally want to study a **population**. You can think of a population as a collection of persons, things, or objects under study. To study the population, we select a **sample**. The idea of **sampling** is to select a portion (or subset) of the larger population and study that portion (the sample) to gain information about the population. Data are the result of sampling from a population.

Because it takes a lot of time and money to examine an entire population, sampling is a very practical technique. If you wished to compute the overall grade point average at your school, it would make sense to select a sample of students who attend the school. The data collected from the sample would be the students' grade point averages. In presidential elections, opinion poll samples of 1,000–2,000 people are taken. The opinion poll is supposed to represent the views of the people in the entire country. Manufacturers of canned carbonated drinks take samples to determine if a 16 ounce can contains 16 ounces of carbonated drink.

From the sample data, we can calculate a statistic. A **statistic** is a number that represents a property of the sample. For example, if we consider one math class to be a sample of the population of all math classes, then the average number of points earned by students in that one math class at the end of the term is an example of a statistic. The statistic is an estimate of a population parameter. A **parameter** is a numerical characteristic of the whole population that can be estimated by a statistic. Since we considered all math classes to be the population, then the average number of points earned per student over all the math classes is an example of a parameter.

One of the main concerns in the field of statistics is how accurately a statistic estimates a parameter. The accuracy really depends on how well the sample represents the population. The sample must contain the characteristics of the population in order to be a **representative sample**. We are interested in both the sample statistic and the population parameter in inferential statistics. In a later chapter, we will use the sample statistic to test the validity of the established population parameter.

A **variable**, usually notated by capital letters such as *X* and *Y*, is a characteristic or measurement that can be determined for each member of a population. Variables may be **numerical** or **categorical**. **Numerical variables** take on values with equal units such as weight in pounds and time in hours. **Categorical variables** place the person or thing into a category. If we let *X* equal the number of points earned by one math student at the end of a term, then *X* is a numerical variable. If we let *Y* be a person's party affiliation, then some examples of *Y* include Republican, Democrat, and Independent. *Y* is a categorical variable. We could do some math with values of *X* (calculate the average number of points earned, for example), but it makes no sense to do math with values of *Y* (calculating an average party affiliation makes no sense).

Data are the actual values of the variable. They may be numbers or they may be words. **Datum** is a single value.

Two words that come up often in statistics are **mean** and **proportion**. If you were to take three exams in your math classes and obtain scores of 86, 75, and 92, you would calculate your mean score by adding the three exam scores and dividing by three (your mean score would be 84.3 to one decimal place). If, in your math class, there are 40 students and 22 are men and 18 are women, then the proportion of men students is $\frac{22}{40}$ and the proportion of women students is $\frac{18}{40}$. Mean and proportion are discussed in more detail in later chapters.

> **NOTE**
>
> The words "**mean**" and "**average**" are often used interchangeably. The substitution of one word for the other is common practice. The technical term is "arithmetic mean," and "average" is technically a center location. However, in practice among non-statisticians, "average" is commonly accepted for "arithmetic mean."

Example 1.1

Determine what the key terms refer to in the following study. We want to know the average (mean) amount of money first year college students spend at ABC College on school supplies that do not include books. We randomly surveyed 100 first year students at the college. Three of those students spent $150, $200, and $225, respectively.
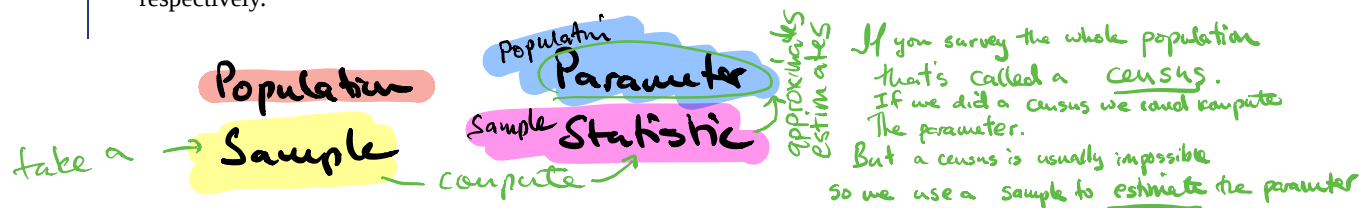
<u>Sample is</u>

those 3 students

<u>Sample statistic is</u>

average amount those 3 students spent

<u>Sample is</u>

100 random students

<u>Sample statistic is</u>

average amount those 100 students spent.

)

[handwritten notes in top margin: Sample / sample statistic / the average amount of money / the]

Solution 1.1

The **population** is all first year students attending ABC College this term.

The **sample** could be all students enrolled in one section of a beginning statistics course at ABC College (although this sample may not represent the entire population).

The **parameter** is the average (mean) amount of money spent (excluding books) by first year college students at ABC College this term.

The **statistic** is the average (mean) amount of money spent (excluding books) by first year college students in the sample.

The **variable** could be the amount of money spent (excluding books) by one first year student. Let X = the amount of money spent (excluding books) by one first year student attending ABC College.

The **data** are the dollar amounts spent by the first year students. Examples of the data are $150, $200, and $225.

Try It Σ

1.1 Determine what the key terms refer to in the following study. We want to know the average (mean) amount of money spent on school uniforms each year by families with children at Knoll Academy. We randomly survey 100 families with children in the school. Three of the families spent $65, $75, and $95, respectively.

Example 1.2

Determine what the key terms refer to in the following study.

A study was conducted at a local college to analyze the average cumulative GPA's of students who graduated last year. Fill in the letter of the phrase that best describes each of the items below.

1. Population ____ 2. Statistic ____ 3. Parameter ____ 4. Sample ____ 5. Variable ____ 6. Data ____

a) all students who attended the college last year
b) the cumulative GPA of one student who graduated from the college last year
c) 3.65, 2.80, 1.50, 3.90
d) a group of students who graduated from the college last year, randomly selected
e) the average cumulative GPA of students who graduated from the college last year
f) all students who graduated from the college last year
g) the average cumulative GPA of students in the study who graduated from the college last year

[handwritten annotations: number; Sample; Populatn; populatn parameter; populatn; since populatn was people, sample will be people; these are not a number so; possibly statistic or parameter; a sample; Populatn & sample are a group of people (or animals or objects or days); Parameter & Statistic are numbers.]